

日本国特許庁
JAPAN PATENT OFFICE

H. Harashima

9/22/03

Q77475

10f1

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日
Date of Application: 2002年10月 1日

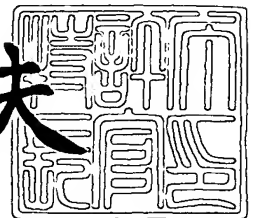
出願番号
Application Number: 特願2002-289071
[ST. 10/C]: [JP2002-289071]

出願人
Applicant(s): 日本電気株式会社

2003年 8月 8日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2003-3063915

【書類名】 特許願

【整理番号】 6700074

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

G11B 20/10

G11B 20/12

【発明者】

【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内

【氏名】 原嶋 一

【特許出願人】

【識別番号】 000004237

【氏名又は名称】 日本電気株式会社

【代理人】

【識別番号】 100082935

【弁理士】

【氏名又は名称】 京本 直樹

【電話番号】 03-3454-1111

【選任した代理人】

【識別番号】 100082924

【弁理士】

【氏名又は名称】 福田 修一

【電話番号】 03-3454-1111

【選任した代理人】

【識別番号】 100085268

【弁理士】

【氏名又は名称】 河合 信明

【電話番号】 03-3454-1111

【手数料の表示】

【予納台帳番号】 008279

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9115699

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 データ書込み、データ書込み読出し装置、データ書込み、データ書込み読出し方法

【特許請求の範囲】

【請求項 1】 上位装置と、記録されるデータが冗長構成となっている第 1 の記憶手段と、前記上位装置からの命令により前記第 1 の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置において、前記制御装置は前記第 1 の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記上位装置から前記アドレスへの書き込み命令のあるデータを第 2 の記憶手段に書込み、前記上位装置に書込み完了報告を行う論理ディスク書込み／読出し手段を備えることを特徴とするデータ書込み装置。

【請求項 2】 前記制御装置は前記アドレスにおける前記冗長破壊が回復されたことを確認する論理ディスク監視手段を更に備え、前記論理ディスク書込み／読出し手段は前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記第 2 の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むことを特徴とする請求項 1 記載のデータ書込み装置。

【請求項 3】 前記論理ディスク監視手段は、一定時間経過したら管理テーブル更新手段に伝えるタイマと、前記第 1 の記憶手段の状態を調べ管理テーブルを更新する前記管理テーブル更新手段と、前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に該回復を前記論理ディスク書込み／読出し手段に報告する書込み可能報告手段とを更に備えることを特徴とする請求項 1，2 記載のデータ書込み装置。

【請求項 4】 前記第 2 の記憶手段は不揮発性の記憶手段であるか、または揮発性であっても独自に電源を持つことを特徴とする請求項 1，2，3 記載のデータ書込み装置。

【請求項 5】 前記第 2 の記憶手段は前記制御装置により書き込まれたデータが前記第 1 の記憶手段に書き込まれるまで前記データを保持することを特徴と

する請求項 1 乃至 4 記載のデータ書込み装置。

【請求項 6】 上位装置と、記録されるデータが冗長構成となっている第 1 の記憶手段と、前記上位装置からの命令により前記第 1 の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置において、前記制御装置は前記第 1 の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記上位装置から前記アドレスへの書き込み命令のあるデータを第 2 の記憶手段に書込み、前記上位装置に書込み完了報告を行い、前記上位装置から前記アドレスへの読出し命令のあるデータがある場合は前記データを第 2 の記憶手段から読出す論理ディスク書込み／読出し手段を備えることを特徴とするデータ書込み読出し装置。

【請求項 7】 前記制御装置は前記アドレスにおける前記冗長破壊が回復されたことを確認する論理ディスク監視手段を更に備え、前記論理ディスク書込み／読出し手段は、前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記第 2 の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むことを特徴とする請求項 6 記載のデータ書込み読出し装置。

【請求項 8】 前記論理ディスク監視手段は、一定時間経過したら管理テーブル更新手段に伝えるタイマと、前記第 1 の記憶手段の状態を調べ管理テーブルを更新する前記管理テーブル更新手段と、前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に該回復を前記論理ディスク書込み／読出し手段に報告する書込み可能報告手段とを更に備えることを特徴とする請求項 6，7 記載のデータ書込み読出し装置。

【請求項 9】 前記第 2 の記憶手段は不揮発性の記憶手段であるか、または揮発性であっても独自に電源を持つことを特徴とする請求項 6，7，8 記載のデータ書込み読出し装置。

【請求項 10】 前記第 2 の記憶手段は前記制御装置により書き込まれたデータを前記制御装置により読出されるまで保持することを特徴とする請求項 6 乃至 9 記載のデータ書込み読出し装置。

【請求項 11】 上位装置と、上位装置から書込み命令を受けたデータ及び冗長データを含み、該冗長データの大きさに相当するデータが破壊されても書込み命令を受けたデータを修復しながら残りのデータにより上位装置からの命令により書込みを行うことのできる第 1 の記憶手段と、前記上位装置からの命令により前記第 1 の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置において、

前記制御装置は前記第 1 の記憶手段のあるアドレスを含む領域においてデータの修理中にエラーにより書込みができない場合、

前記上位装置から前記アドレスへの書き込み命令のあるデータを第 2 の記憶手段に書込み、前記上位装置に書込み完了報告を行う論理ディスク書込み／読出し手段を備えることを特徴とするデータ書込み装置。

【請求項 12】 前記制御装置は前記データの修復が完了したことを確認する論理ディスク監視手段を更に備え、

前記論理ディスク書込み／読出し手段は前記論理ディスク監視手段が前記データの修復が完了したことを確認すると、前記第 2 の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むことを特徴とする請求項 11 記載のデータ書込み装置。

【請求項 13】 前記論理ディスク監視手段は、一定時間経過したら管理テーブル更新手段に伝えるタイマと、

前記第 1 の記憶手段の状態を調べ管理テーブルを更新する前記管理テーブル更新手段と、

前記管理テーブルにおいて前記アドレスにおける前記データの修復が完了した場合に該回復を前記論理ディスク書込み／読出し手段に報告する書込み可能報告手段とを更に備えることを特徴とする請求項 11, 12 記載のデータ書込み装置。

【請求項 14】 前記第 2 の記憶手段は不揮発性の記憶手段であるか、または揮発性であっても独自に電源を持つことを特徴とする請求項 11, 12, 13 記載のデータ書込み装置。

【請求項 15】 前記第 2 の記憶手段は前記制御装置により書き込まれたデータが前記第 1 の記憶手段に書き込まれるまで前記データを保持することを特徴

とする請求項 11 乃至 14 記載のデータ書込み装置。

【請求項 16】 上位装置と、記録されるデータが冗長構成となっている第 1 の記憶手段と、前記上位装置からの命令により前記第 1 の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置のデータ書込み方法において、
前記制御装置は論理ディスク書込み／読出し手段を備え、
前記論理ディスク書込み／読出し手段が前記第 1 の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記上位装置から前記アドレスへの書き込み命令のあるデータを第 2 の記憶手段に書込むステップと、
前記上位装置に書込み完了報告を行うステップを備えることを特徴とするデータ書込み方法。

【請求項 17】 前記制御装置は論理ディスク監視手段を更に備え、
前記論理ディスク監視手段が前記アドレスにおける前記冗長破壊が回復されたことを確認するステップと、
前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記論理ディスク書込み／読出し手段が前記第 2 の記憶手段に書き込まれたデータを読出すステップと、
前記記憶手段の前記アドレスに前記データを書き込むステップとを更に備えることを特徴とする請求項 16 記載のデータ書込み方法。

【請求項 18】 前記論理ディスク監視手段はタイマと管理テーブル更新手段と書込み可能報告手段とを更に備え、
一定時間経過したら前記タイマが管理テーブル更新手段に伝えるステップと、
前記管理テーブル更新手段が前記第 1 の記憶手段の状態を調べ管理テーブルを更新するステップと、
書込み可能報告手段が前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に該回復を前記論理ディスク書込み／読出し手段に報告するステップとを更に備えることを特徴とする請求項 16, 17 記載のデータ書込み方法。

【請求項 19】 上位装置と、記録されるデータが冗長構成となっている第

1 の記憶手段と、前記上位装置からの命令により前記第 1 の記憶手段にデータの書き込みを行う制御装置とを備えたデータ書き込み装置において、前記制御装置は論理ディスク書き込み／読出し手段を更に備え、前記第 1 の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記論理ディスク書き込み／読出し手段が、前記上位装置から前記アドレスへの書き込み命令のあるデータを第 2 の記憶手段に書き込むステップと、前記論理ディスク書き込み／読出し手段が、前記上位装置に書き込み完了報告を行うステップと、前記上位装置から前記アドレスへの読出し命令のあるデータがある場合は、前記論理ディスク書き込み／読出し手段が、前記データを第 2 の記憶手段から読出するステップとを備えることを特徴とするデータ書き込み読出し方法。

【請求項 20】 前記制御装置は前記アドレスにおける前記冗長破壊が回復されたことを確認する論理ディスク監視手段を更に備え、前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記論理ディスク書き込み／読出し手段が前記第 2 の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むステップを更に備えることを特徴とする請求項 19 記載のデータ書き込み読出し方法。

【請求項 21】 前記論理ディスク監視手段は、タイマと管理テーブル更新手段と書き込み可能報告手段とを備え、一定時間経過したら前期タイマが管理テーブル更新手段に伝えるステップと、前記管理テーブル更新手段が前記第 1 の記憶手段の状態を調べ管理テーブルを更新するステップと、前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に、前期書き込み可能報告手段が、該回復を前記論理ディスク書き込み／読出し手段に報告するステップとを更に備えることを特徴とする請求項 19, 20 記載のデータ書き込み読出し方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明はデータ書込み、書込み読出し装置、データ書込み、書込み読出し方法に関し、特に冗長構成を備えた記憶装置において冗長破壊が起こった場合のデータ書込み、読出し装置、データ書込み、書込み読出し方法に関する。

【0002】

【従来の技術】

従来のデータ書込み装置として例えば、ディスクアレイ装置がある。

【0003】

図10は従来のディスクアレイ装置であり、上位装置100とコントローラ101と論理ディスク120の構成を示すブロック図である。

【0004】

上位装置100はデータの書込み／読出しを論理ディスク120から行うために、コントローラ1101に対しデータの書込み／読出し命令を送信し、コントローラ1101に該当データを転送し、また読出しの場合はコントローラ1101から該当データを受信する。

【0005】

コントローラ1101は論理ディスク監視手段111と論理ディスク書込み／読出し手段112とを備える。

【0006】

論理ディスク監視手段111は管理テーブル更新手段150とタイマ114を備えている。タイマ114は電源が入っている状態では定期的に管理テーブル更新手段150に一定時間が経過したことを伝える。タイマ114から一定時間が経過したことを伝えられると管理テーブル更新手段150は論理ディスク120の各論理アドレスの状態を監視する。その結果そして、管理テーブル更新手段150は上位装置とは非同期に論理ディスク120の状態を調べる。

【0007】

論理ディスク書込み／読出し手段112は上位装置100からデータの書込み命令とデータを受信すると、上位装置100の指定する論理ディスク120の該当論理アドレスにデータの書込みを行う。書込みができない場合は規定回数リトライ動作を行う。規定回数リトライ動作を行っても書込みを完了することができ

ない場合は上位装置 1 0 0 に書込みエラー報告を行う。書込みを完了すると上位装置 1 0 0 に対して書込み正常終了報告を行う。書込みを完了することができなかった場合は上位装置 1 0 0 に対し書込みエラー報告を行う。

【0 0 0 8】

論理ディスク書込み／読出し手段 1 1 2 は上位装置 1 0 0 からデータの読出し命令を受信すると、上位装置 1 0 0 の指定する論理ディスク 1 2 0 の該当論理アドレスからデータの読出しを行う。読出しができない場合は規定回数リトライ動作を行う。規定回数リトライ動作を行っても読出しを完了することができない場合は上位装置 1 0 0 に読出しエラー報告を行う。読出しを完了すると上位装置 1 0 0 に対して読出し正常終了報告を行う。読出しを完了することができなかった場合は上位装置 1 0 0 に対し読出しエラー報告を行う。

【0 0 0 9】

論理ディスク 1 2 0 は複数の HDD を備え、ここでは HDD 1 3 0 A ~ HDD 1 3 0 E の 5 台を備えている例を示す。コントローラ 1 1 0 1 からデータの書込み命令を受けると、コントローラ 1 1 0 1 に指定された論理アドレスに冗長性を持ってデータは書込まれる。ここでは論理アドレス 1 4 0 A に書込まれる例を示す。またコントローラ 1 1 0 1 からデータの読出し命令を受けると、コントローラ 1 1 0 1 に指定された論理アドレスからデータは読出される。ここでは論理アドレス 1 4 0 A から読出される例を示す。

【0 0 1 0】

ここで、論理ディスク 1 2 0 は HDD 1 3 0 E における論理アドレス 1 4 0 A に媒体エラーがあり、かつ HDD 1 3 0 A はディスク障害により HDD 交換後、データ修復を行っている途中の状態である。このため論理アドレス 1 4 0 A は二重障害状態にありアクセス不可とされるため書込みを行うことができない。また、同様に論理アドレス 1 4 0 A からデータの読出しを行うことはできない。

【0 0 1 1】

逆に論理アドレス 1 4 0 B は HDD 1 3 0 E において媒体エラーはなく HDD 1 3 0 A のみデータ修復中であるので、書込み、読出しとも行うことができる。

【0 0 1 2】

図 11 は上位装置 100 とコントローラ 1101 と論理ディスク 120 とを備える従来のディスクアレイ装置の動作を示し、上位装置 100 が論理ディスク 120 に書込みを行うときの動作を示したフローチャートである。

【0013】

上位装置 100 がコントローラ 1101 に論理ディスク 120 への書込みデータを送信すると（ステップ P1）、コントローラ 1101 内の論理ディスク書込み／読出し手段 112 が書込みデータを受信する（ステップ P2）。次に、論理ディスク書込み／読出し手段 112 は上位装置 100 が指定する論理ディスク 120 の論理アドレス 140A へデータの書込みを行う。このとき書込みができない場合は規定回数リトライ動作を行う。初回から規定回数までのリトライ動作の間に論理アドレス 140A へ書込みを完了したら（ステップ P3, Y）、論理ディスク書込み／読出し手段 112 は上位装置 100 に書込み正常終了を報告する（ステップ P4）。規定回数までのリトライ動作の間に論理アドレス 140A へ書込みを完了できなかった場合は（ステップ P3, N）は、上位装置 100 に書込みエラーを報告する（ステップ P5）。

【0014】

図 12 は上位装置 100 とコントローラ 1101 と論理ディスク 120 とを備える従来のディスクアレイ装置の動作を示し、上位装置 100 が論理ディスク 120 から読出しを行うときの動作を示したフローチャートである。

【0015】

上位装置 100 がコントローラ 1101 に論理ディスク 120 からデータを読出す、読出し命令を送信すると（ステップ Q1）、コントローラ 1101 内の論理ディスク書込み／読出し手段 112 が読出し命令を受信する（ステップ Q2）。次に、論理ディスク書込み／読出し手段 112 は上位装置 100 が指定する論理ディスク 120 の論理アドレス 140A からデータの読出しを行う。このとき読出しができない場合は規定回数リトライ動作を行う。初回から規定回数までのリトライ動作の間に論理アドレス 140A のデータ読出しを完了したら（ステップ Q3, Y）、論理ディスク書込み／読出し手段 112 は上位装置 100 に読出したデータを転送する（ステップ Q4）。規定回数までのリトライ動作の間に論

理アドレス 1 4 0 A からの読出し完了できなかった場合は（ステップ Q 3, N）は、上位装置 1 0 0 に読出しエラーを報告する（ステップ Q 5）。

【 0 0 1 6 】

次に、論理ディスク監視手段 1 1 1 の動作を説明する。図 1 3 は論理ディスク監視手段 1 1 1 の動作を示したフローチャートである。コントローラ 1 1 0 1 に電源が入ると、タイマ 1 1 4 は時間計測を開始し一定時間が経過すると管理テーブル更新手段 1 5 0 に伝える（ステップ R 1）。管理テーブル更新手段 1 5 0 は論理ディスク 1 2 0 の状態を調べ、その状態を基に管理テーブル 1 5 1 を更新する（ステップ R 2）。タイマ 1 1 4 はコントローラ 1 1 0 1 に電源が入っている限り動作を継続する。論理ディスク監視手段 1 1 1 の上記動作は上位装置 1 0 0 の動作とは非同期に行われる。

【 0 0 1 7 】

次に、図 1 0 に召すように論理ディスク 1 2 0 が HDD 1 3 0 A と HDD 1 3 0 E において二重障害状態にある場合の図チのフローチャートにおける従来のディスクアレイ装置の動作を説明する。論理ディスク 1 2 0 は HDD 1 3 0 E における論理アドレス 1 4 0 A 部に媒体エラーがあり、かつ HDD 1 3 0 A はディスク障害により HDD 交換後、データ修復を行っている途中の状態である。このように論理アドレス 1 4 0 A は二重障害状態にあるため、書込みを行うことができない。そのため、規定回数リトライ動作を行っても論理アドレス 1 4 0 A への書込みを完了することができない（ステップ P 3, N）。そして、論理ディスク書込み／読出し手段 1 1 2 は上位装置 1 0 0 に書込みエラーを報告をすることになる（ステップ P 5）。

【 0 0 1 8 】

このようにディスクアレイ装置は HDD の一時的二重障害の発生時には、上位装置から HDD へ一部のデータ書込みができないため、またその結果該当データの読出しができないため、上位装置へ書込みエラー報告或いは読出しエラー報告を行う I/O エラーが発生する可能性が高くなる（図 1 1、ステップ P 5）。そして、I/O エラーが発生した場合、原因の特定及びデータの再書込み等の付加作業が必要となり、通常業務に支障をきたすことになる。

【0019】

このような二重障害対策として従来のディスクアレイ装置では二重障害からの回復のためにパリティデータを二重化しているものがある（例えば、特許文献1参照）。

【0020】**【特許文献1】**

特開2000-39970号公報

また、一般のライトキャッシュ方式を採用したディスクアレイ装置に於いては、上位装置からのデータを一時的にライトキャッシュに格納した時点で上位装置へ書込み正常終了報告を行うことにより、I/Oエラーの発生する可能性を低下させることが可能な場合も考えられる。

【0021】

図10、図11を用いて説明すると、図10における論理ディスク書込み／読出し手段112がライトキャッシュを備え、図11におけるステップP2において、コントローラ1101内の論理ディスク書込み／読出し手段112が書込みデータを受信すると、上位装置100に書込み正常終了報告を行うことに相当する。そして二重障害発生中の上位装置からの該当データの読出しはライトキャッシュから行い、障害発生中のHDDの修復処理の完了を待ってライトキャッシュから該当データをHDDに書込むことによって二重障害発生時のデータアクセスの救済を行い、I/Oエラーの発生する可能性を低下させること可能な場合もあると考えられる。しかしながら、次のような問題点が考えられる。

【0022】**【発明が解決しようとする課題】**

第1の問題点は、ライトキャッシュを採用した救済方法では、ライトキャッシュが常に有効状態である必要がある点である。通常このようなディスクアレイ装置ではライトキャッシュのON/OFFが可能であり、仮に設定上はON（有効）状態の場合でもディスクアレイ装置内の各リソースの動作状態により障害発生時などにはライトキャッシュが強制的にOFF（無効）状態とされることが多い。この場合、ライトキャッシュにデータを一時的に保存することができないため

、HDDの一時的二重障害発生時にはHDDへの書込み読出しができず、I/Oエラーとなる。

【0023】

第2の問題点は、ライトキャッシュが有効状態であった場合においてもデータ書込みができず、I/Oエラーとなる点である。論理ディスクが修復中のケースに於いて上位装置からの書込みデータをいったんライトキャッシュに保存した後、順次論理ディスクへ書込みを開始する。HDDの一時的二重障害発生時にデータ書込みがエラーとなった場合、規定回数のリトライ動作中に修復対象のHDDの該当アドレスの修復処理が完了すればリトライ動作によるデータの書込みが可能となる。しかし、リトライ動作中に修復対象のHDDの該当アドレスの修復処理が完了できない場合はライトキャッシュからのデータ書込みができずI/Oエラーとなる。また、論理ディスクが縮退中の場合は該当アドレス部で常にHDDの二重障害状態となるためリトライ動作での書込みも不可であることから上記と同様にライトキャッシュからのデータ書込みができずI/Oエラーとなる。

【0024】

ライトキャッシュを採用したディスクアレイ装置ではコントローラが、上位装置からのデータの書込みを完了すると、コントローラが上位装置へ書込み正常終了報告を行う。そして正常終了報告後、ライトキャッシュからHDDへ書込みが行えない場合であっても上位装置へエラー報告を行わない装置の場合は、上位装置から見るとI/Oエラーが生じていないことになる。しかしながら、HDDの該当アドレスに書込みが行えていないことに変わりはなく、従って、該当アドレスのデータを読出すこともできない。

【0025】

本発明の目的は、二重障害の発生時に上位装置としてはデータの書込み、読出しを正常終了させることが可能となり、障害の発生アドレスの修復中も該当アドレスのデータを読出すことが可能となり、障害の発生アドレスの修復が完了した時点でHDDに書込みを完了することを可能とするデータ書込み装置、データ書込み読出し装置を提供することにある。

【0026】

【課題を解決するための手段】

本発明のデータ書込み装置は上位装置と、記録されるデータが冗長構成となっている第1の記憶手段と、前記上位装置からの命令により前記第1の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置において、前記制御装置は前記第1の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記上位装置から前記アドレスへの書き込み命令のあるデータを第2の記憶手段に書込み、前記上位装置に書込み完了報告を行う論理ディスク書込み／読出し手段を備えることを特徴とする。

【0027】

請求項2記載のデータ書込み装置は前記制御装置は前記アドレスにおける前記冗長破壊が回復されたことを確認する論理ディスク監視手段を更に備え、前記論理ディスク書込み／読出し手段は前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記第2の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むことを特徴とする。

【0028】

請求項3記載のデータ書込み装置は前記論理ディスク監視手段は、一定時間経過したら管理テーブル更新手段に伝えるタイマと、前記第1の記憶手段の状態を調べ管理テーブルを更新する前記管理テーブル更新手段と、前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に該回復を前記論理ディスク書込み／読出し手段に報告する書込み可能報告手段とを更に備えることを特徴とする。

【0029】

請求項4記載のデータ書込み装置は、前記第2の記憶手段は不揮発性の記憶手段であるか、または揮発性であっても独自に電源を持つことを特徴とする。

【0030】

請求項5記載のデータ書込み装置は前記第2の記憶手段は前記制御装置により書き込まれたデータが前記第1の記憶手段に書き込まれるまで前記データを保持することを特徴とする。

【0031】

請求項 6 記載のデータ書込み読出し装置は、上位装置と、記録されるデータが冗長構成となっている第 1 の記憶手段と、前記上位装置からの命令により前記第 1 の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置において、前記制御装置は前記第 1 の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記上位装置から前記アドレスへの書き込み命令のあるデータを第 2 の記憶手段に書込み、前記上位装置に書込み完了報告を行い、前記上位装置から前記アドレスへの読出し命令のあるデータがある場合は前記データを第 2 の記憶手段から読出す論理ディスク書込み／読出し手段を備えることを特徴とする。

【0032】

請求項 7 記載のデータ書込み読出し装置は、前記制御装置は前記アドレスにおける前記冗長破壊が回復されたことを確認する論理ディスク監視手段を更に備え、前記論理ディスク書込み／読出し手段は、前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記第 2 の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むことを特徴とする。

【0033】

請求項 8 記載のデータ書込み読出し装置は、前記論理ディスク監視手段は、一定時間経過したら管理テーブル更新手段に伝えるタイマと、前記第 1 の記憶手段の状態を調べ管理テーブルを更新する前記管理テーブル更新手段と、前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に該回復を前記論理ディスク書込み／読出し手段に報告する書込み可能報告手段とを更に備えることを特徴とする。

【0034】

請求項 9 記載のデータ書込み読出し装置は、前記第 2 の記憶手段は不揮発性の記憶手段であるか、または揮発性であっても独自に電源を持つことを特徴とする。

【0035】

請求項 10 記載のデータ書込み読出し装置は、前記第 2 の記憶手段は前記制御

装置により書き込まれたデータを前記制御装置により読出されるまで保持することを特徴とする。

【0036】

請求項11記載のデータ書き込み装置は、上位装置と、上位装置から書き込み命令を受けたデータ及び冗長データを含み、該冗長データの大きさに相当するデータが破壊されても書き込み命令を受けたデータを修復しながら残りのデータにより上位装置からの命令により書き込みを行うことのできる第1の記憶手段と、前記上位装置からの命令により前記第1の記憶手段にデータの書き込みを行う制御装置とを備えたデータ書き込み装置において、前記制御装置は前記第1の記憶手段のあるアドレスを含む領域においてデータの修理中にエラーにより書き込みができない場合、前記上位装置から前記アドレスへの書き込み命令のあるデータを第2の記憶手段に書き込み、前記上位装置に書き込み完了報告を行う論理ディスク書き込み／読出し手段を備えることを特徴とする。

【0037】

請求項12記載のデータ書き込み装置は、前記制御装置は前記データの修復が完了したことを確認する論理ディスク監視手段を更に備え、前記論理ディスク書き込み／読出し手段は前記論理ディスク監視手段が前記データの修復が完了したことを確認すると、前記第2の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むことを特徴とする。

【0038】

請求項13記載のデータ書き込み装置は、前記論理ディスク監視手段が、一定時間経過したら管理テーブル更新手段に伝えるタイマと、前記第1の記憶手段の状態を調べ管理テーブルを更新する前記管理テーブル更新手段と、前記管理テーブルにおいて前記アドレスにおける前記データの修復が完了した場合に該回復を前記論理ディスク書き込み／読出し手段に報告する書き込み可能報告手段とを更に備えることを特徴とする。

【0039】

請求項14記載のデータ書き込み装置は、前記第2の記憶手段は不揮発性の記憶手段であるか、または揮発性であっても独自に電源を持つことを特徴とする。

【0040】

請求項15記載のデータ書込み装置は、前記第2の記憶手段は前記制御装置により書き込まれたデータが前記第1の記憶手段に書き込まれるまで前記データを保持することを特徴とする。

【0041】

請求項16記載のデータ書込み方法は、上位装置と、記録されるデータが冗長構成となっている第1の記憶手段と、前記上位装置からの命令により前記第1の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置のデータ書込み方法において、前記制御装置は論理ディスク書込み／読出し手段を備え、前記論理ディスク書込み／読出し手段が前記第1の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記上位装置から前記アドレスへの書き込み命令のあるデータを第2の記憶手段に書込むステップと、前記上位装置に書込み完了報告を行うステップを備えることを特徴とする。

【0042】

請求項17記載のデータ書込み方法は、前記制御装置は論理ディスク監視手段を更に備え、前記論理ディスク監視手段が前記アドレスにおける前記冗長破壊が回復されたことを確認するステップと、前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記論理ディスク書込み／読出し手段が前記第2の記憶手段に書き込まれたデータを読出すステップと、前記記憶手段の前記アドレスに前記データを書き込むステップとを更に備えることを特徴とする。

【0043】

請求項18記載のデータ書込み方法は、前記論理ディスク監視手段はタイマと管理テーブル更新手段と書込み可能報告手段とを更に備え、一定時間経過したら前記タイマが管理テーブル更新手段に伝えるステップと、前記管理テーブル更新手段が前記第1の記憶手段の状態を調べ管理テーブルを更新するステップと、書込み可能報告手段が前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に該回復を前記論理ディスク書込み／読出し手段に報告するステップとを更に備えることを特徴とする。

【0044】

請求項 19 記載のデータ書込み読出し方法は、上位装置と、記録されるデータが冗長構成となっている第 1 の記憶手段と、前記上位装置からの命令により前記第 1 の記憶手段にデータの書込みを行う制御装置とを備えたデータ書込み装置において、前記制御装置は論理ディスク書込み／読出し手段を更に備え、前記第 1 の記憶手段のあるアドレスにおいて冗長破壊が発生した場合に、前記論理ディスク書込み／読出し手段が、前記上位装置から前記アドレスへの書き込み命令のあるデータを第 2 の記憶手段に書込むステップと、前記論理ディスク書込み／読出し手段が、前記上位装置に書込み完了報告を行うステップと、前記上位装置から前記アドレスへの読出し命令のあるデータがある場合は、前記論理ディスク書込み／読出し手段が、前記データを第 2 の記憶手段から読出するステップとを備えることを特徴とする。

【0045】

請求項 20 記載のデータ書込み読出し方法は、前記制御装置は前記アドレスにおける前記冗長破壊が回復されたことを確認する論理ディスク監視手段を更に備え、前記論理ディスク監視手段が前記冗長破壊が回復されたことを確認すると、前記論理ディスク書込み／読出し手段が前記第 2 の記憶手段に書き込まれたデータを読出し、前記記憶手段の前記アドレスに書き込むステップを更に備えることを特徴とする。

【0046】

請求項 21 記載のデータ書込み読出し方法は、前記論理ディスク監視手段は、タイマと管理テーブル更新手段と書込み可能報告手段とを備え、一定時間経過したら前期タイマが管理テーブル更新手段に伝えるステップと、前記管理テーブル更新手段が前記第 1 の記憶手段の状態を調べ管理テーブルを更新するステップと、前記管理テーブルにおいて前記アドレスにおける前記冗長破壊が回復された場合に、前期書込み可能報告手段が、該回復を前記論理ディスク書込み／読出し手段に報告するステップとを更に備えることを特徴とする。

【0047】

【発明の実施の形態】

次に、本発明の第 1 の実施の形態について図を参照して詳細に説明する

図1は本発明の第1の実施の形態における上位装置100とコントローラ110と論理ディスク120の構成を示すブロック図である。図11に示された従来のディスクアレイ装置と比較して、コントローラ110にメモリ113がある点、論理ディスク監視手段111に書込み可能報告手段160がある点で異なる。

【0048】

上位装置100はデータの書込み／読出しを論理ディスク120に対して行うために、コントローラ110に対しデータの書込み／読出し命令を送信し、コントローラ110に該当データを転送し、また読出しの場合はコントローラ110から該当データを受信する。またコントローラ110から書込み正常終了報告、書込みエラー報告、読出しエラー報告を受信する。書込み正常終了報告を受信した場合は、上位装置100として該当データの書込みを正常終了する。

【0049】

コントローラ110は論理ディスク監視手段111と論理ディスク書込み／読出し手段112とメモリ113とを備える。

【0050】

論理ディスク監視手段111はタイマ114と管理テーブル更新手段150と書込み可能報告手段160とを備えている。タイマ114は電源が入っている状態では定期的にタイマ114が管理テーブル更新手段150に一定時間が経過したことを伝える。そして、管理テーブル更新手段150は上位装置とは非同期に論理ディスク120の状態を調べる。管理テーブル更新手段150は内部に管理テーブル151（図5参照）を備え、論理ディスク120の状態に応じて順次その管理テーブル151を更新する。

【0051】

書込み可能報告手段160は論理ディスク書込み／読出し手段112から書込みを行うことができなかった論理アドレスを知らされると、管理テーブル151を常に参照する。そして、論理ディスク監視手段111が書込みを行おうとして行えなかった論理アドレス、即ち本実施の形態では論理アドレス140AがReassign OK（代替処理OK、二重障害が解除されたことを示す）の状態となると、論理ディスク書込み／読出し手段112に書込みを行おうとして行えなかった論

理アドレスにおいて書込みが可能になったことを伝える。尚、管理テーブル 151 については後述する。

【0052】

論理ディスク書込み／読出し手段 112 は上位装置 100 からデータの書込み命令とデータを受信すると、上位装置 100 の指定する論理ディスク 120 の該当論理アドレスにデータの書込みを行う。書込みを行うことができない場合は規定回数リトライ動作を行う。書込みを完了すると上位装置 100 に対して書込み正常終了報告を行う。規定回数リトライ動作を行っても書込みを完了することができない場合は該当データをメモリ 113 に書込み、書込み可能報告手段 160 に書込みを行えなかった論理アドレスを伝え、上位装置 100 に書込み正常終了報告を行う。そして、書込み可能報告手段 160 より書込みを行おうとした論理アドレス、即ち本実施の形態では論理アドレス 140 A が Reassign OK (代替処理 OK、二重障害が解除されたことを示す) の状態となったことを知らされると、メモリ 113 内のデータを論理ディスク 120 内の書込みを行おうとした論理アドレス 140 A に書込む。

【0053】

論理ディスク書込み／読出し手段 112 は上位装置 100 からデータの読出し命令を受信すると、上位装置 100 の指定する論理ディスク 120 の該当論理アドレス、本実施の形態では論理アドレス 140 A からデータの読出しを行う。読出しを行うことができない場合は規定回数リトライ動作を行う。読出しを完了すると上位装置 100 へ読出したデータを転送する。規定回数リトライ動作を行っても読出しを完了することができない場合は該当データがメモリ 113 に記録されているかどうか判断する。メモリ 113 に該当データが記録されていた場合は読出し、上位装置 100 へ該当データを転送する。メモリ 113 に記録されていない場合は、上位装置 100 に読出しエラー報告を行う。

【0054】

メモリ 113 は、論理ディスク書込み／読出し手段 112 から転送されてきたデータを書込み、論理ディスク書込み／読出し手段 112 から読出し命令のあるデータを読出し論理ディスク書込み／読出し手段 112 に送出する。ここでメモ

リ 1 1 3 は HDD 1 3 0 E の媒体欠陥箇所（特定アドレス部（ここでは論理アドレス 1 4 0 A）でのメディアエラー）への書込み不可データを一時的に保存する部分であるため、バッファサイズとしては数十 K B から数百 K B 程度の容量でよい。また、メモリ 1 1 3 は書き込まれたデータが論理ディスク書込み／読出し手段 1 1 2 から読出され、論理ディスク 1 2 0 に書込まれるまで保持する。

【0055】

論理ディスク 1 2 0 は複数の単体 HDD を備え、ここでは HDD 1 3 0 A ~ HDD 1 3 0 E の 5 台を備えている例を示す。そして論理ディスク 1 2 0 は冗長構成となっていて、RAID の構成としては RAID 1 / 3 / 5 などが使用される。コントローラ 1 1 0 からデータの書込み命令を受けると、コントローラ 1 1 0 に指定された論理アドレスに冗長性を持ってデータは書込まれる。ここでは論理アドレス 1 4 0 A に書込まれる例を示す。また論理ディスク 1 2 0 はコントローラ 1 1 0 からデータの読出し命令を受けると、コントローラ 1 1 0 に指定された論理アドレスからデータは読出される。ここでは論理アドレス 1 4 0 A から読出される例を示す。通常時、論理ディスク 1 2 0 は上述のとおり冗長構成となっているため、論理ディスク 1 2 0 内の HDD 1 3 0 A ~ HDD 1 3 0 E のうち任意の HDD が 1 台故障した場合（冗長内故障時）でもコントローラ 1 1 0 からのデータを書込んだり、他のディスクのパリティーデータを元にデータを生成することによりデータを読出してコントローラ 1 1 0 に転送することができる。論理アドレス 1 4 0 B はこの冗長内故障の状態にあるアドレスである。

【0056】

ここで、図 1 に示した論理ディスク 1 2 0 は HDD 1 3 0 A がディスク障害により HDD 交換後データ修復を行っている途中の状態であり、HDD 1 3 0 E における論理アドレス 1 4 0 A 部には媒体エラーが発生している。そのため論理アドレス 1 4 0 A においては HDD 2 台が故障している二重障害状態であり、データを読出そうとしても他のディスクのパリティーデータを元にデータを生成することができない冗長破壊（冗長外故障）の状態である。そのため、論理アドレス 1 4 0 A へアクセスが不可となるため書込みを行うことができない。また、同様に論理アドレス 1 4 0 A からデータの読出しを行うことはできない。逆に論理ア

ドレス 140B は HDD 130E において媒体エラーはなく HDD 130A のみデータ修復中であるので、書込み、読出しとも行うことができる。

【0057】

データの修復は上位装置 100 とは非同期に論理ディスク 120 を構成する他の HDD 130B ~ HDD 130E のパリティデータを基に HDD 130A へのデータの修復が実施される。尚、RAID 構成によってはこのパリティデータを特定の 1 台の HDD に持つタイプもある。

【0058】

次に図 1 に示した上位装置 100 とコントローラ 110 と論理ディスク 120 の動作を図 1、図 2 を参照して詳細に説明する。

【0059】

図 2 は本発明の第 1 の実施の形態における上位装置 100 とコントローラ 110 と論理ディスク 120 の動作を示し、上位装置 100 が論理ディスク 120 に書込みを行うときの動作を示したフローチャートである。従来技術の上位装置 100 とコントローラ 110 と論理ディスク 120 の動作を示したフローチャートである図 11 とステップ A5 からステップ A10 までの処理が異なる。尚、本発明第 1 の実施の形態において示す処理動作は論理ディスク書込み／読出し手段 112 と論理ディスク論理ディスク監視手段 111 が該プログラムを記録した記録媒体（磁気ディスク、磁気テープ、半導体メモリ、CD-ROM、DVD (Digital Versatile Disk) 等の光ディスク）から読出し装置を介して読出すか、もしくはサーバ等から通信媒体を介してダウンロードすることでコントローラ 110 のプロセッサ（図示せず）の記録媒体（ハードディスク等、図示せず）にインストールし、該プロセッサのメモリ（図示せず）に該プログラムを読出し、実行することで実現される。

【0060】

上位装置 100 がコントローラ 110 に論理ディスク 120 への書込みデータを送信すると（ステップ A1）、コントローラ 110 内の論理ディスク書込み／読出し手段 112 が書込みデータを受信する（ステップ A2）。次に、論理ディスク書込み／読出し手段 112 は上位装置 100 が指定する論理ディスク 120

の論理アドレス 140 A へデータの書込みを行う。このとき書込みができない場合は規定回数リトライ動作を行う。初回から規定回数までのリトライ動作の間に論理アドレス 140 A へ書込みを行うことができた場合は（ステップ A 3， Y）、論理ディスク書込み／読出し手段 112 は上位装置 100 に書込み正常終了を報告し、この時点で上位装置 100 からの該当データの書込みは正常終了する（ステップ A 4）。

【0061】

次に、図 1 に示すように論理ディスク 120 が HDD 130 A と HDD 130 E の二重障害状態にある場合を説明する。論理ディスク 120 は HDD 130 E における論理アドレス 140 A 部に媒体エラーがあり、かつ HDD 130 A はディスク障害により HDD 交換後、データ修復を行っている途中の状態である。このため論理アドレス 140 A は自アドレスにおいてデータ修復、代替処理が行われるまでは一時的に二重障害状態にあるため、書込みを行うことができない。そのため、規定回数リトライ動作を行っても論理アドレス 140 A への書込みを完了することができず、その場合は書込み不可能と判断する（ステップ A 3， N）。

【0062】

論理ディスク書込み／読出し手段 112 は論理ディスク 120 の論理アドレス 140 A に代わってメモリ 113 にデータを書込み、書込みを行えなかった論理アドレス 140 A を書込み可能報告手段 160 に伝える（ステップ A 5）。論理ディスク書込み／読出し手段 112 はメモリ 113 への書込みを終えると、上位装置 100 に書込み正常終了を報告する。この時点で上位装置 100 からの該当データの書き込みは正常終了する（ステップ A 6）。この後、論理ディスク書込み／読出し手段 112 に対して書込み可能報告手段 160 からステップ A 3 において書込みを行おうとして行えなかった論理アドレス 140 A について書込みが可能となったという報告があると（ステップ A 7）、論理ディスク書込み／読出し手段 112 はメモリ 113 内のデータを論理ディスク 120 内の論理アドレス 140 A へ書込む（ステップ A 8）。ここで、メモリ 113 は保持している論理アドレス 140 A に書き込まれるべきであったデータを少なくとも論理ディスク

書込み／読出し手段 112 がメモリ 113 内のデータを読出し論理ディスク 120 の論理アドレス 140 A へ書き込むまでは保持している。

【0063】

論理ディスク書込み／読出し手段 112 がメモリ 113 内のデータを論理ディスク 120 内の論理アドレス 140 A へ書込むときに書込みを完了できなかった場合（ステップ A9，N）、論理ディスク書込み／読出し手段 112 は上位装置 100 にエラー報告をする（ステップ A10）。書込みを完了できた場合は（ステップ A9，Y）、はそのまま処理を終了する。

【0064】

ここで図 3 の説明をする。図 3 は、上位装置 100 とは非同期に行っていた論理ディスク 120 における HDD 130 A において論理アドレス 140 A 部までデータの修復及び代替処理が完了し、論理アドレス 140 A 部の HDD 130 A と HDD 130 E における一時的な二重障害が解除されたことを示した図である。前述のように論理アドレス 140 A は修復することはできないので実際の修復処理はスキップされており、該当データの書込みがされているわけではない。しかし、媒体エラー発生箇所以外の場所へデータ領域を移動させる代替処理が完了している。即ち、論理アドレス 140 A の修復が完了し Reassign OK（代替処理 OK、後述）の状態である。そのため一時的な二重障害が解除されたことになる。少なくとも本実施の形態においては論理アドレス 140 A のデータの書込みに関してはデータの書込みが可能になったため、二重障害は確実に解除されたことになる。

【0065】

図 4 は上位装置 100 とコントローラ 110 と論理ディスク 120 の動作を示し、上位装置 100 が論理ディスク 120 から読出しを行うときの動作を示したフローチャートである。従来技術の上位装置 100 とコントローラ 110 と論理ディスク 120 の動作を示したフローチャートである図 12 とステップ B5 からステップ B7 までの処理が異なる。

【0066】

上位装置 100 がコントローラ 110 に論理ディスク 120 からデータを読出

す、読出し命令を送信すると（ステップB1）、コントローラ110内の論理ディスク書込み／読出し手段112が読出し命令を受信する（ステップB2）。次に、論理ディスク書込み／読出し手段112は上位装置100が指定する論理ディスク120の論理アドレス140Aからのデータ読出しを行う。このとき読出しができない場合は規定回数リトライ動作を行う。初回から規定回数までのリトライ動作の間に論理アドレス140Aから読出しを完了することができた場合は（ステップB3、Y）、論理ディスク書込み／読出し手段112は上位装置100に読出したデータを転送する（ステップB4）。

【0067】

初回から規定回数までのリトライ動作の間に論理アドレス140Aから読出しを完了することができなかった場合は（ステップB3、N）、論理ディスク書込み／読出し手段112は該当データがメモリ113に記録されているかどうか判断し（ステップB5）、メモリ113に記録されている場合は（ステップB5、Y）、論理ディスク書込み／読出し手段112はメモリ113からデータを読出し（ステップB6）、論理ディスク書込み／読出し手段112は読出したデータを上位装置100へ転送する（ステップB7）。該当データがメモリ113に記録されていなかった場合は（ステップB5、N）、論理ディスク書込み／読出し手段112は上位装置100に読出しエラー報告をする（ステップB8）。

【0068】

次に、管理テーブル151の説明をする。図5は管理テーブル更新手段150の内部にある管理テーブル151の全体図を示した図である。管理テーブル151は論理ディスク120の各論理アドレスの状態を示し、論理ディスク監視手段111が論理ディスク120から得た情報により順次書き換えられ、また書込み可能報告手段160が論理ディスク書込み／読出し手段112に論理ディスク120の状態を報告するときは、管理テーブル151を参照する。

【0069】

テーブルの横の項目は論理ディスク120の論理アドレスを示す。ここでは複数の論理アドレスがある中、140A（論理アドレス140A）、140B（論理アドレス140B）を代表して示している。論理アドレス140Aは図1に示

したとおり HDD 130A がデータ修復中であり、HDD 130E の論理アドレス 140A に媒体エラーがあるため二重障害の状態にある（図 5 では図示せず）。

【0070】

テーブルの縦の項目はそれぞれの論理ディスクの状態を示す項目であり、1. Repair はパリティデータを元に修復 (Repair) が完了したかどうかを示し、完了した場合は、Repair OK と記載され、完了していない場合は Repair NG、修復 (Repair) を行わず飛ばした場合は、Repair skip と記載される。

【0071】

2. Reassign の有無はディスクにおいて媒体エラーのため使用不可能である場合に、代替領域を物理的にディスクの別の場所に用意したか、どうかを示す。Reassign OK の場合は代替処理が終了し、Reassign NG の場合は代替処理が終了していないことを示す。ディスクにおいて媒体エラーが発生しておらず、そもそも代替処理の必要がない場合はこの項目は空欄となる。

【0072】

3. Read/Write は読出し、書込みが可能かどうかを示す。読出し書き込みが可能の場合 Read OK, Write OK と記載され、読出し書き込みが不可能な場合 Read NG, Write NG と記載される。次に、図 5 における各論理アドレスにおける具体的な例を示す。

【0073】

図 6 は図 5 における論理アドレス 140A に関する情報の詳細を示した図である。まず 1 番目の 1. Repair の項目に Repair skip と記載されている。これは論理アドレス 140A は二重障害状態にあったためパリティデータを元に修復 (Repair) を行うことができず論理アドレス 140A は修復処理を飛ばした (skip) したことを示す。

【0074】

次に 2 番目として 2. Reassign の有無の項目に Reassign OK と記載されている。これは HDD 130E の論理アドレス 140A の領域は媒体エラーのため使用不可能であるため、代替領域を物理的にディスクの別の場所に用意したことを示

す。この代替領域も論理的には論理アドレス 140A として扱われるため、論理アドレス 140A においても冗長性を持ってデータを書き込むことが可能となる（本実施の形態の場合 5 台の HDD 全てにデータを書き込むことが可能となる。）。そのため、Reassign OKにより、論理アドレス 140A におけるデータの二重障害は解除されたことを示す。

【0075】

次に 3 番目として 3. Read/Write の項目に READ NG / WRITE OK と記載されている。これは上位装置 100 からのアクセスに対して論理アドレス 140A の領域は、Reassign の処理により書込みは冗長性をもって行うことができるが、それまでの二重障害のためデータを回復させることができなかったためにデータの保証を行うことができず、読出しは不可能であることを示す。

【0076】

図 7 は図 5 における論理アドレス 140B に関する情報の詳細を示した図である。まず 1 番目として 1. Repair の項目に Repair OK と記載されている。これは論理アドレス 140B においてパリティデータを元に修復 (Repair) が正常に終了したことを示す。

【0077】

次に 2 番目として 2. Reassign の有無の項目に Reassign の記載はない。論理アドレス 140B は代替領域が必要でないためである。

【0078】

次に 3 番目として 3. Read/Write の項目に READ OK / WRITE OK と記載されている。これは上位装置 100 からのアクセスに対して論理アドレス 140B の領域は、書込みも読出しも可能であることを示す。

【0079】

次に、論理ディスク監視手段 111 の動作を説明する。論理ディスク監視手段 111 は従来の技術と比較して書込み可能報告手段 160 を有する点が異なるので、書込み可能報告手段 160 の動作を説明する。図 8 は書込み可能報告手段 160 の動作を示したフローチャートである。書込み可能報告手段 160 は論理ディスク書込み／読出し手段 112 から論理ディスク 120 へ書込みを行えなかつ

た論理アドレス 140 A を知らされると（ステップ C 1）、管理テーブル 151 を参照し、論理アドレス 140 A において Reassign OK（代替処理 OK、書込み可能）となっているか調べることで該当アドレスが書き込み可能であるか調べ、Reassign OK（代替処理 OK、書込み可能）となっている場合は（ステップ C 2、Y）、書込み可能報告手段 160 が論理ディスク書込み／読出し手段 112 に書込み可能報告を行う（ステップ C 3）。Reassign OK（代替処理 OK、書込み可能）となっていない場合は（ステップ C 2、N）、再びステップ C 2 の動作を行う。

【0080】

ここで、このステップ C 3 の処理が図 2 におけるステップ A 7 と対応する。また、従来の技術における従来の論理ディスク監視手段 111 の動作を示したフローチャートである図 13 において説明したとおり、管理テーブル 151 は一定間隔で更新されている。そのため、Reassign の項目が更新されると、書込み可能報告手段 160 によって該当アドレスが書き込み可能であることが認識される。

【0081】

このように本発明の第 1 の実施の形態では、二重障害の発生時にも上位装置としてはデータの書込み、読出しを正常終了させることができ、I/O エラーを発生しない。その理由は、該当する論理アドレスへのデータの書き込みが不可能な場合は、メモリへ書込み（即ち退避させ）、上位装置へ書込み正常終了報告を行うためである。また、該当する論理アドレスからのデータの読出しが不可能な場合は、メモリから退避中のデータを読出す構成としたためである。

【0082】

また、本発明の第 1 の実施の形態では、二重障害発生時においてもデータ書込み命令のあるデータの書込みを行うことができる。その理由は、該当する論理アドレスへのデータ書込みが不可能な場合はメモリへ書込み、論理アドレスの修復が完了後、メモリに退避中のデータを HDD に書込むためである。

【0083】

また、障害の発生した論理アドレスの修復完了後すぐに HDD にデータの書込みを行うこともできる。その理由は、メモリへデータの書込みを行った後上位装

置とは非同期に定期的に論理ディスク内の該当する論理アドレスが書込み可能であるかどうか調べるためである。

【0084】

次に、本発明の第2の実施の形態について図を参照して詳細に説明する。本発明の第1の実施の形態ではHDD130Aの修復動作中のためHDD二重障害が発生している場合のデータアクセスに関して救済する方法を説明したが、本発明の第2の実施の形態ではHDD130Aが縮退動作中のためHDD二重障害が発生している場合のデータアクセスに関して救済する方法を説明する。

【0085】

図9は本発明の第2の実施の形態における上位装置100とコントローラ110と論理ディスク120の構成を示すブロック図である。図1の本発明の第1の実施の形態における構成を示すブロック図と比較してメモリ113が不揮発性メモリ1131である、もしくは括弧書きで示したメモリ113とバッテリー1132である点が異なる。また、HDD130Aが修復動作中であつたものが、縮退動作中である点が異なる。

【0086】

次に図9に示した上位装置100とコントローラ110と論理ディスク120の動作を図9、図2を参照して詳細に説明する。本発明の第2の実施の形態の動作は第1の実施の形態の動作とほぼ同一であるため、第2の実施の形態の動作が第1の実施の形態の動作と異なる点のみを説明する。

【0087】

本発明の第1の実施の形態と同様図2において、HDD130Aが縮退動作中に上位装置100がコントローラ110に論理ディスク120への書込みデータを送信すると（ステップA1）、その後の処理は本発明の第1の実施の形態において説明した動作と同じ動作を行う。しかしながら縮退中のHDD130Aがすぐに保守作業により修復されとは限らないため、ディスクアレイ装置の電源断等によりメモリ内のデータが損失しないよう不揮発性メモリ1131を用いる、もしくはバッテリー1132によってメモリバックアップされたメモリ113を用いる。

【0088】

本発明の第2の実施の形態では、HDDの縮退動作により二重障害が発生している場合においても上位装置としてはデータの書込み、読出しを正常終了させることが可能となり、縮退動作中も該当アドレスのデータを読出すことが可能となり、障害の発生アドレスの修復が完了した時点でHDDに書込みを完了することができる。その理由は、メモリ内のデータが損失しないよう不揮発性メモリを用いる、もしくはバッテリーによってメモリバックアップされたメモリを用いているためである。

【0089】**【発明の効果】**

本発明は以下の効果により、HDDの一時的な二重障害発生時に於いても上位装置からのデータアクセスが可能となり、論理ディスクの冗長内故障時の信頼性を向上させることができる。

【0090】

本発明の第1の効果は、二重障害の発生時にも上位装置としてはデータの書込み、読出しを正常終了させることができ、I/Oエラーを発生しないことである。その理由は、該当する論理アドレスへのデータの書き込みが不可能な場合は、メモリへ書込み（即ち退避させ）、上位装置へ書込み正常終了報告を行うためである。また、該当する論理アドレスからのデータの読出しが不可能な場合は、メモリから退避中のデータを読出す構成としたためである。

【0091】

第2の効果は、二重障害発生時においてもデータ書込み命令のあるデータの書込みを行うことができることである。その理由は、該当する論理アドレスへのデータ書込みが不可能な場合はメモリへ書込む。そして、論理アドレスの修復が完了後、メモリに退避中のデータをHDDに書込むためである。

【0092】

第3の効果は、障害の発生した論理アドレスの修復完了後すぐにHDDにデータの書込みを行うことができることである。その理由は、メモリへデータの書込みを行った後上位装置とは非同期に定期的に論理ディスク内の該当する論理アド

レスが書込み可能であるかどうか調べ、書込みであるとすぐにメモリに書き込まれているデータをHDDに書込むためである。

【0093】

第4の効果は、HDDの縮退動作により二重障害が発生している場合においても上位装置としてはデータの書込み、読出しを正常終了させることが可能となり、障害の発生アドレスの修復中も該当アドレスのデータを読出すことが可能となり、障害の発生アドレスの修復が完了した時点でHDDに書込みを完了することができることである。その理由は、メモリ内のデータが損失しないよう不揮発性メモリを用いる、もしくはバッテリーによってメモリバックアップされたメモリを用いているためである。

【図面の簡単な説明】

【図1】本発明の第1の実施の形態における上位装置100とコントローラ110と論理ディスク120の構成を示すブロック図

【図2】本発明の第1の実施の形態における上位装置100とコントローラ110と論理ディスク120の動作を示し、上位装置100が論理ディスク120に書込みを行うときの動作を示したフローチャート

【図3】本発明の第1の実施の形態における上位装置100とコントローラ110と論理ディスク120の構成を示すブロック図であり、上位装置100とは非同期に行っていた論理ディスク120におけるHDD130Aにおいて論理アドレス140A部までデータの修復が完了し、論理アドレス140A部のHDD130AとHDD130Eにおける二重障害が解除されたことを示した図

【図4】本発明の第1の実施の形態における上位装置100とコントローラ110と論理ディスク120の動作を示し、上位装置100が論理ディスク120から読出しを行うときの動作を示したフローチャート

【図5】管理テーブル更新手段150の全体図を示した図

【図6】論理アドレス140Aに関する情報の詳細を示した図

【図7】論理アドレス140Bに関する情報の詳細を示した図

【図8】論理ディスク監視手段111の動作を示したフローチャート

【図9】本発明の第2の実施の形態における上位装置100とコントローラ

110と論理ディスク120の構成を示すブロック図であり、HDD130Aが縮退動作中であることを示した図

【図10】従来の上位装置100とコントローラ110と論理ディスク120の構成を示すブロック図

【図11】従来の上位装置100とコントローラ110と論理ディスク120とを備える従来のディスクアレイ装置の動作を示し、上位装置100が論理ディスク120に書き込みを行うときの動作を示したフローチャート

【図12】従来の上位装置100とコントローラ110と論理ディスク120とを備える従来のディスクアレイ装置の動作を示し、上位装置100が論理ディスク120から読出しを行うときの動作を示したフローチャート

【図13】従来の論理ディスク監視手段111の動作を示したフローチャート

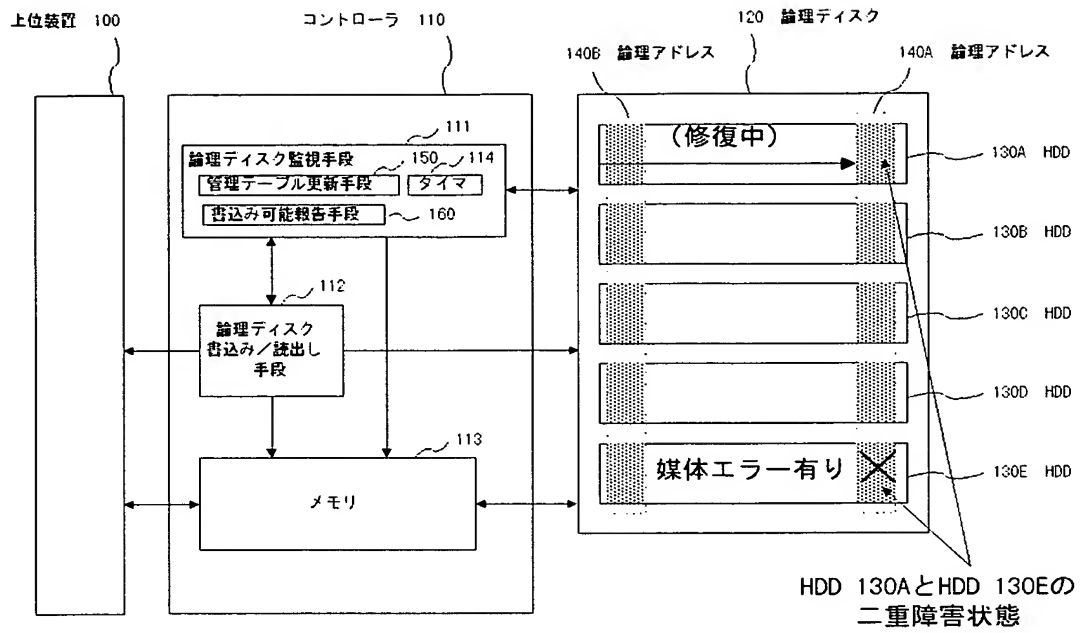
【符号の説明】

100	上位装置
110	コントローラ
111	論理ディスク監視手段
112	論理ディスク書き込み／読出し手段
113	メモリ
114	タイマ
120	論理ディスク
130A	HDD
130B	HDD
130C	HDD
130D	HDD
130E	HDD
140A	論理アドレス
140B	論理アドレス
150	管理テーブル更新手段
151	管理テーブル

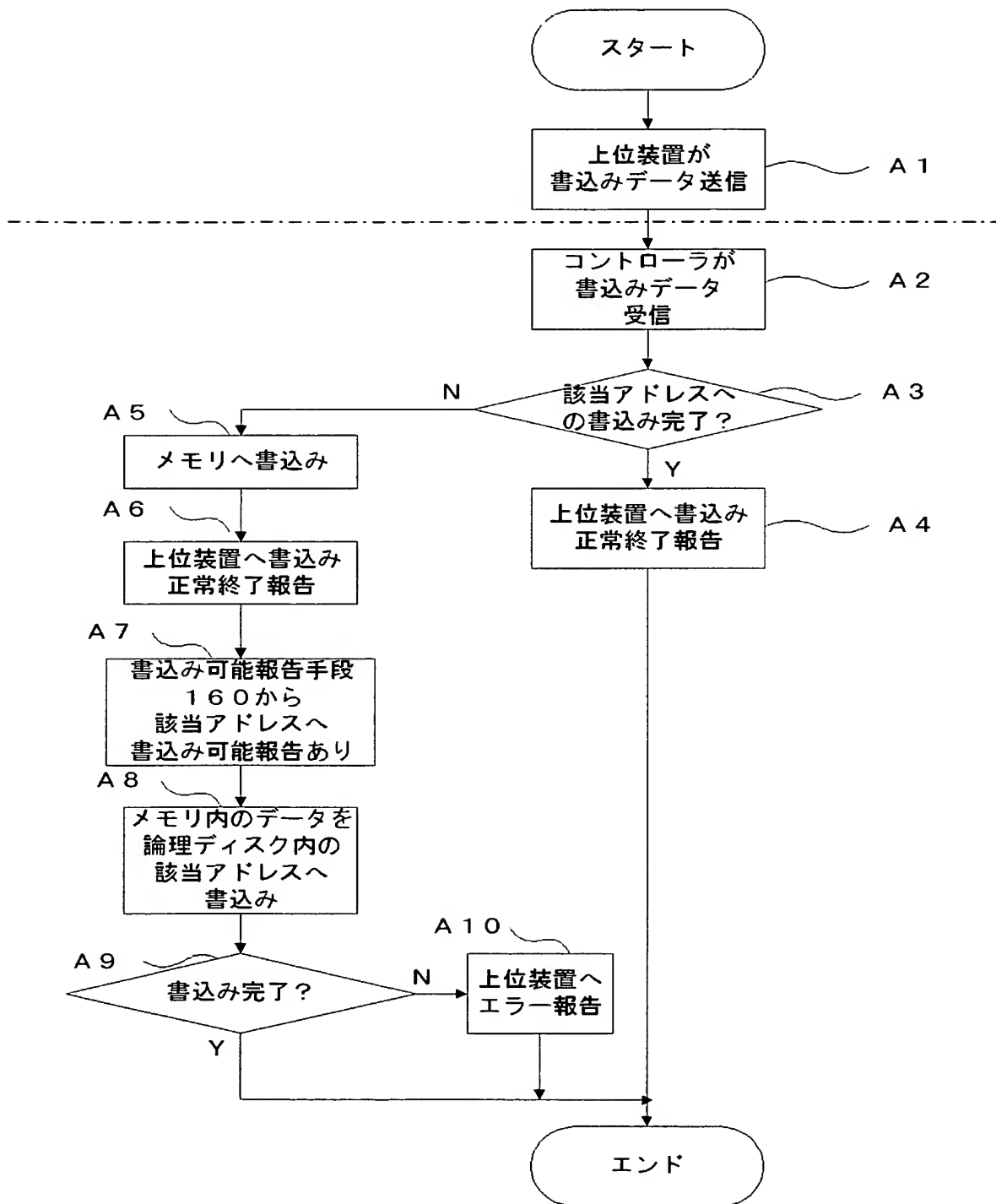
1 6 0 , 書込み可能報告手段
1 1 0 1 コントローラ
1 1 3 1 不揮発性メモリ
1 1 3 2 バッテリ

【書類名】 図面

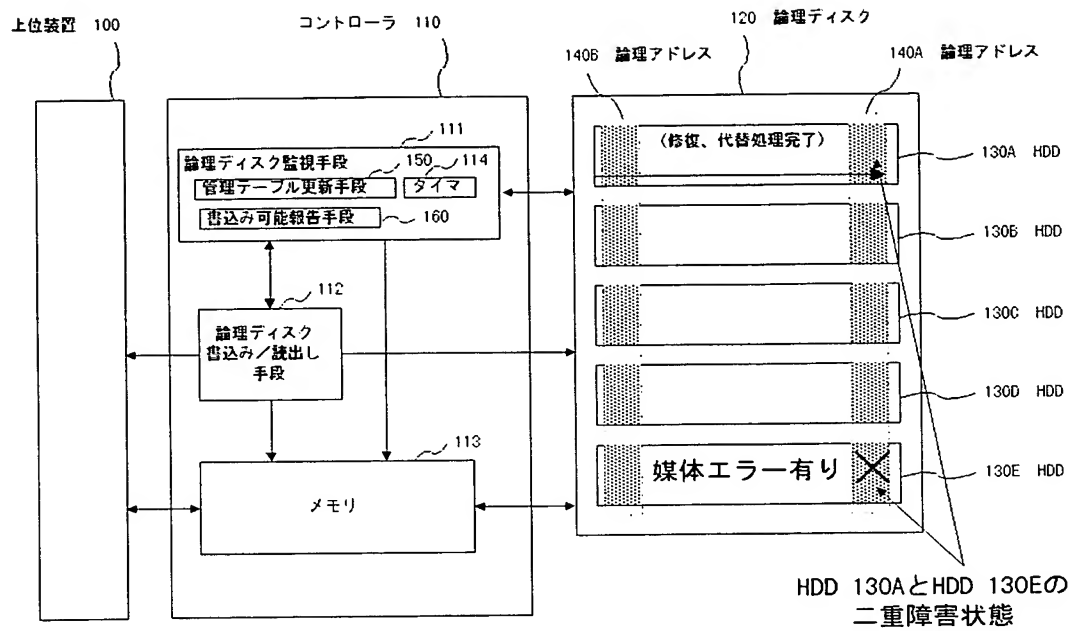
【図 1】



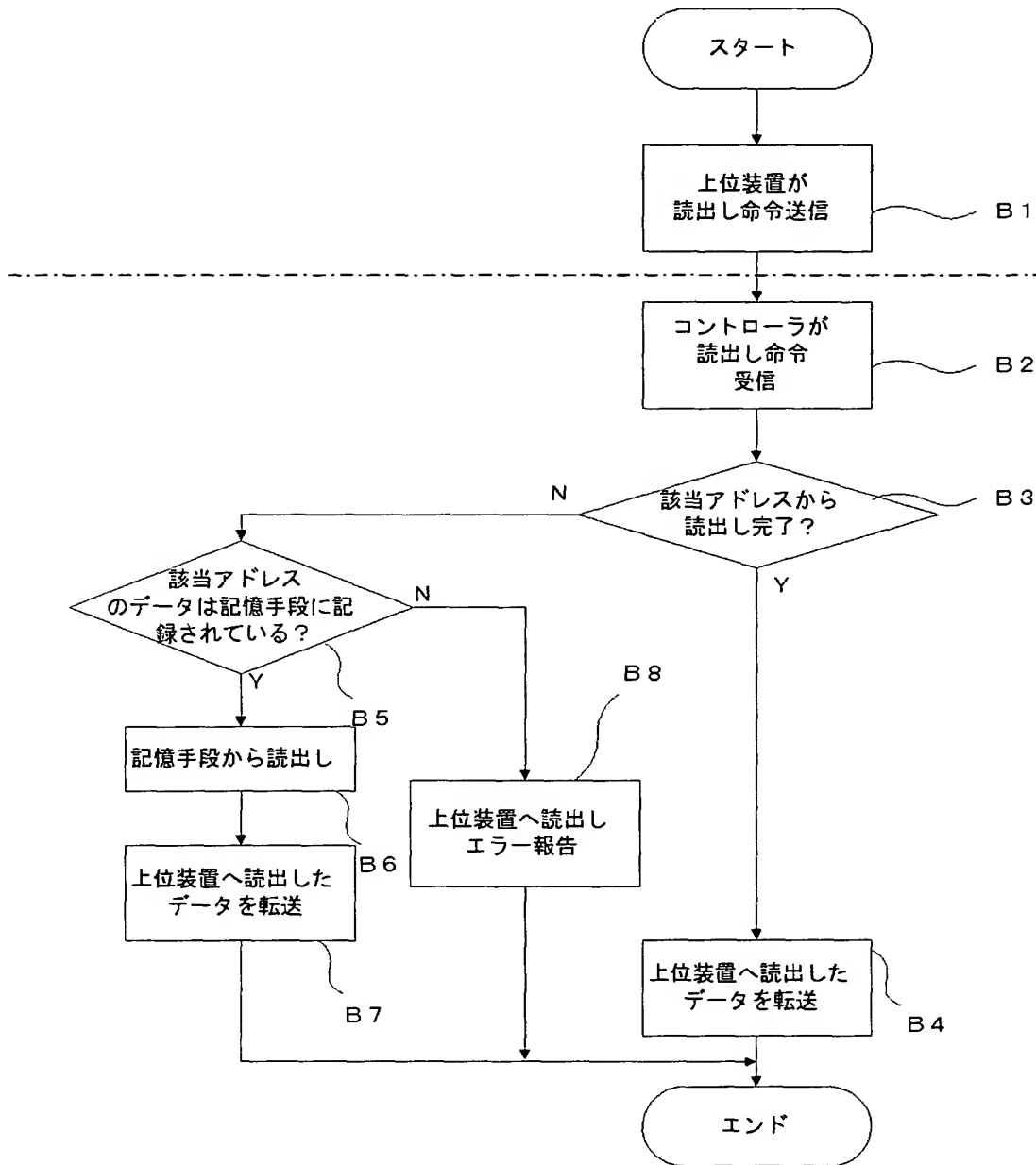
【図 2】



【図 3】



【図 4】



【図 5】

管理テーブル 151

論理アドレス	...	140B	...	140A	...
1. Repair					
2. Reassign の有無					
3.	Read				
	Write				

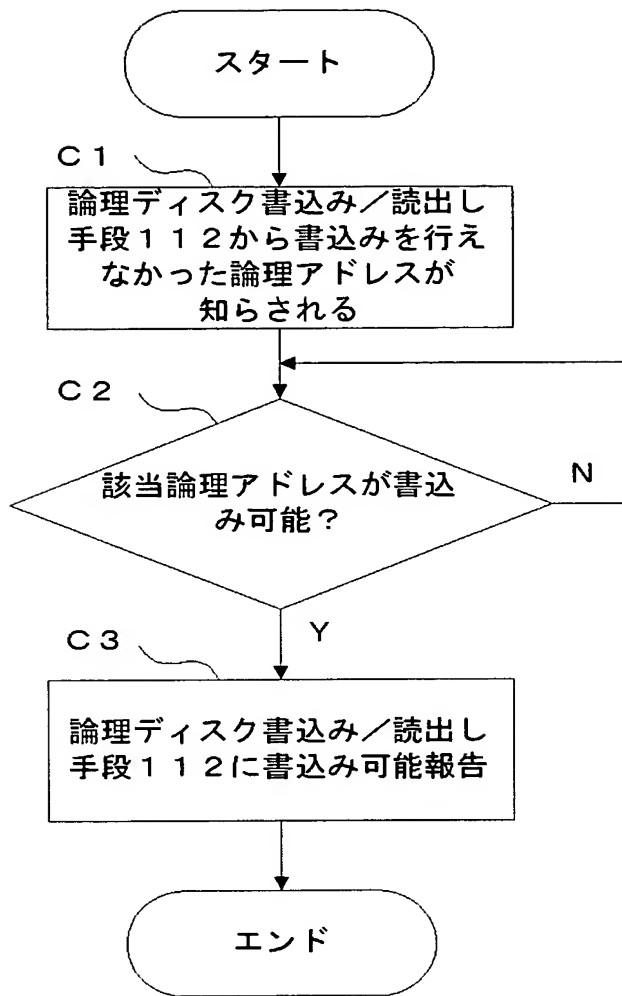
【図 6】

論理アドレス	140A	
1. Repair	Repairskip	
2. Reassign の有無	Reassign OK	
3. Read / Write	READ NG	WRITE OK

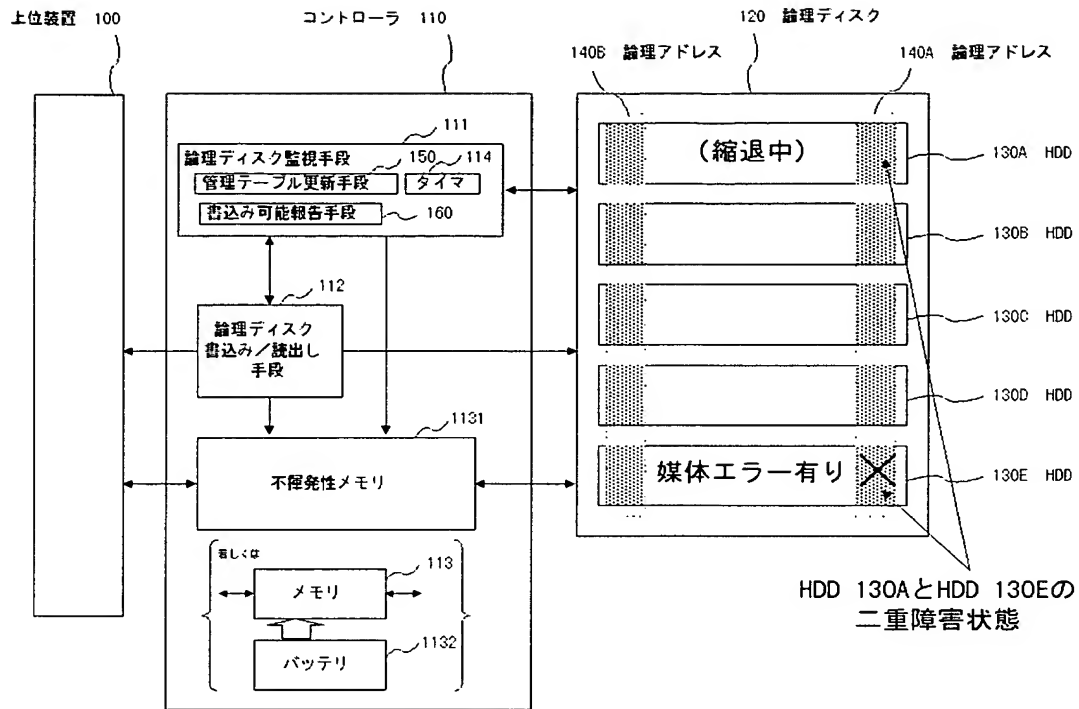
【図 7】

論理アドレス	1 4 0 B	
1. Repair	Repair OK	
2. Reassign の有無		
3. Read / Write	READ OK	WRITE OK

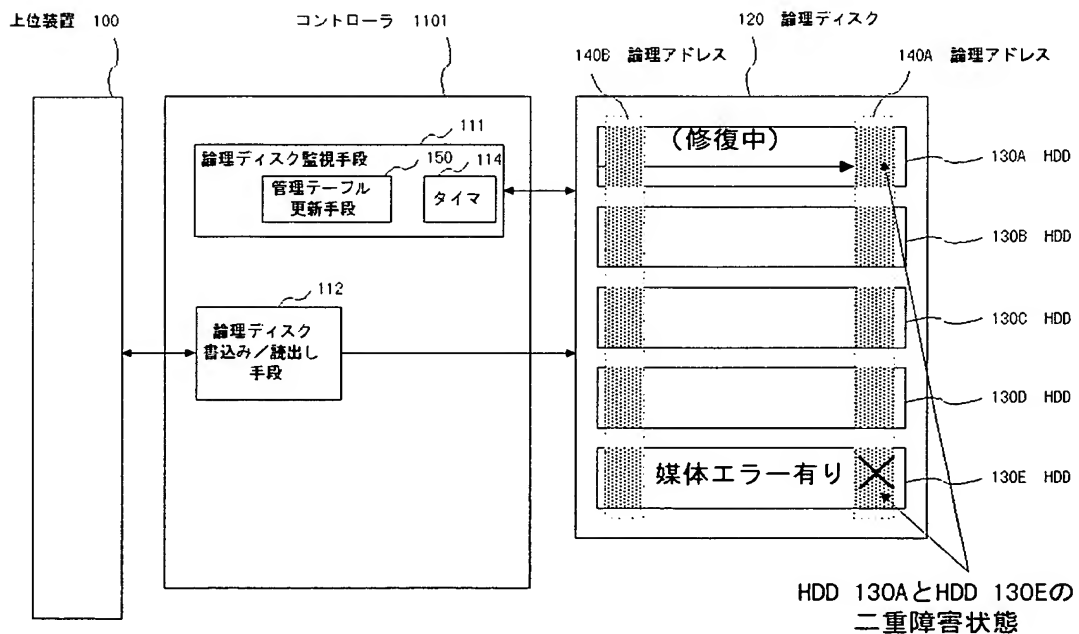
【図 8】



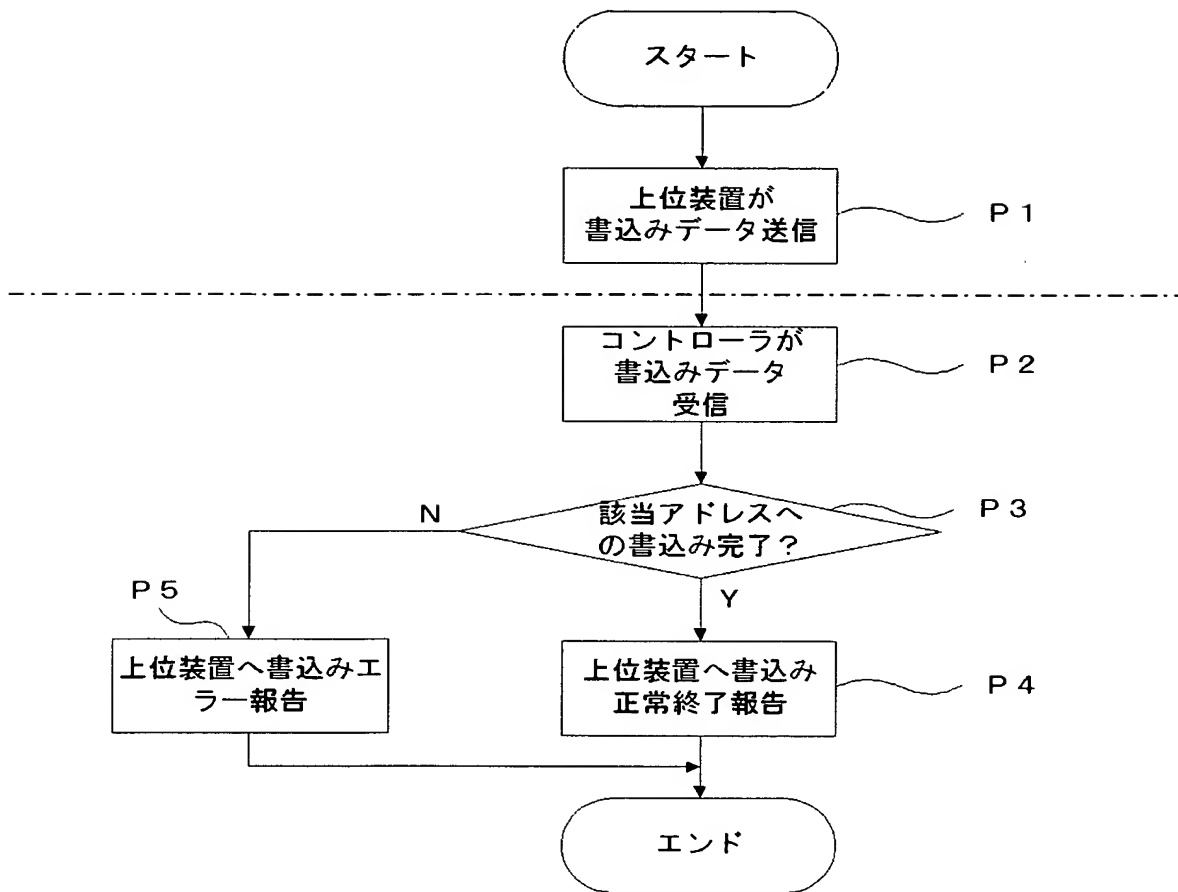
【図 9】



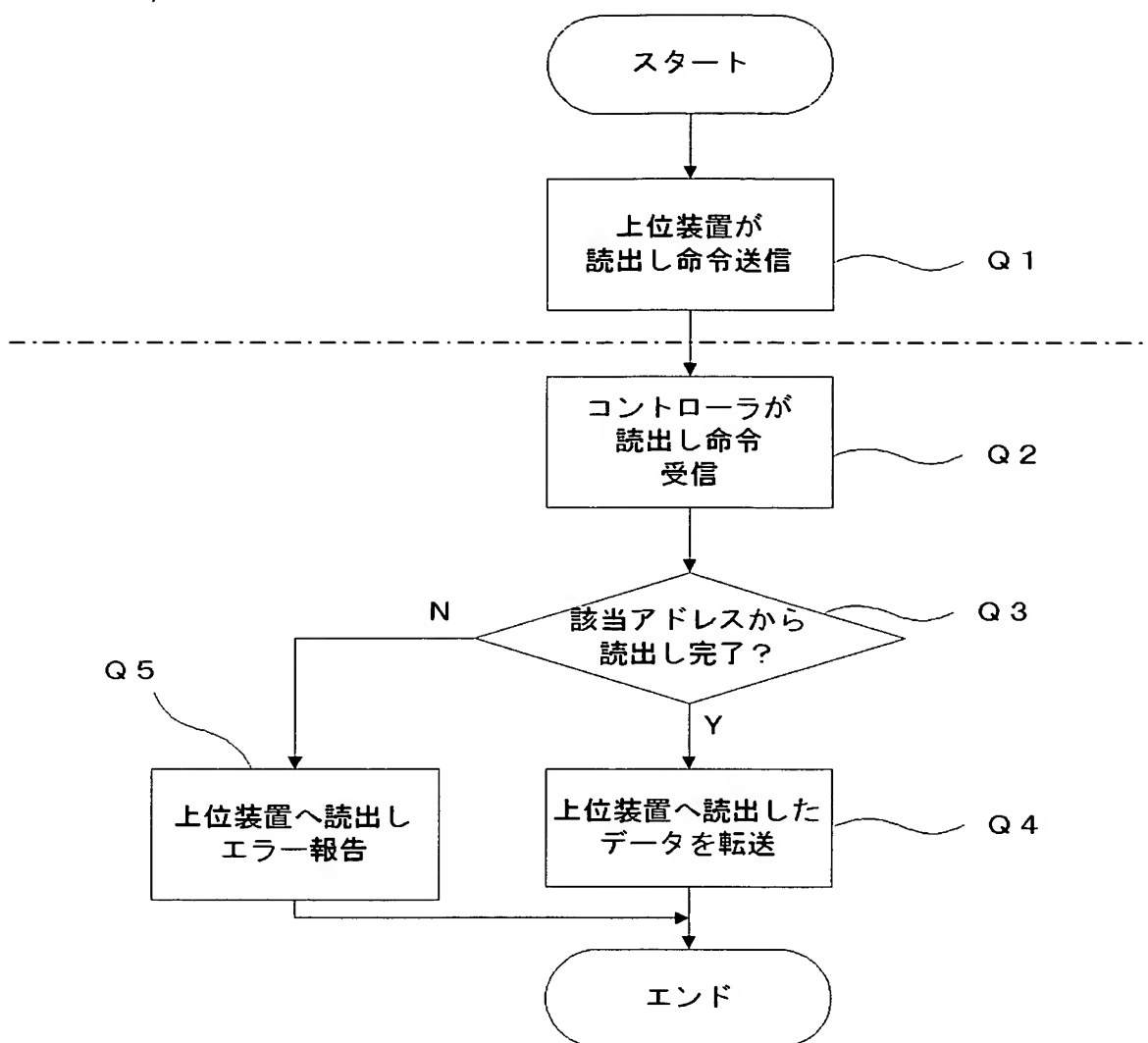
【図 10】



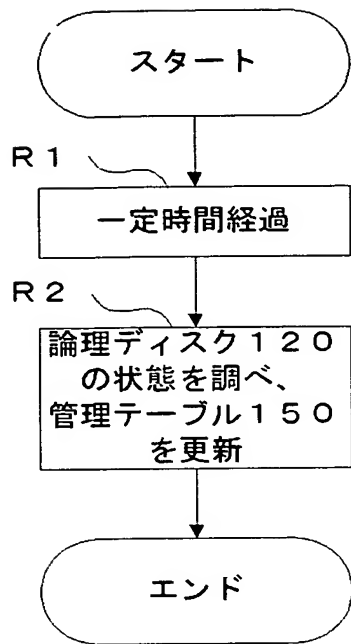
【図 11】



【図 12】



【図 13】



【書類名】 要約書

【要約】

【課題】 二重障害の発生時に上位装置はデータの書込み、読出しを正常終了させ、障害の発生アドレスの修復中も該当アドレスのデータを読出し、障害の発生アドレスの修復が完了した時点でHDDに書込みを完了することを可能とするデータ書込み装置、データ書込み読出し装置を提供する。

【解決手段】 論理ディスク書込み／読出し手段112は論理ディスク120の論理アドレス140Aに対し二重障害により書込みを行えなかった場合、メモリ113に該当データを書込み、上位装置100には書込み正常終了報告を行う。上位装置100の該当データの読出しはメモリ113から行う。論理アドレス140Aが書込み可能になったことを書込み可能報告手段160から報告された論理ディスク書込み／読出し手段112は、メモリ113内の該当データを論理ディスク120の論理アドレス140Aに書込む。

【選択図】 図1

認定・付加情報

特許出願の番号	特願 2 0 0 2 - 2 8 9 0 7 1
受付番号	5 0 2 0 1 4 7 8 2 2 5
書類名	特許願
担当官	第七担当上席 0 0 9 6
作成日	平成 1 4 年 1 0 月 2 日

< 認定情報・付加情報 >

【提出日】	平成14年10月 1日
-------	-------------

次頁無

特願 2 0 0 2 - 2 8 9 0 7 1

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 4 2 3 7]

1 . 変 更 年 月 日

1 9 9 0 年 8 月 2 9 日

[変 更 理 由]

新 規 登 録

住 所

東 京 都 港 区 芝 五 丁 目 7 番 1 号

氏 名

日 本 電 気 株 式 会 社